Intelligent Automation And Soft Computing, 2019

Vol. 25, no. 2, 249–257
https://doi.org/10.31209/2018.100000010

Image Classification using Optimized MKL for sSPM

Lu Wu, Quan Liu, Ping Lou

School of Information Engineering, Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Wuhan University of Technology, Wuhan, Hubei, China

ABSTRACT

The scheme of spatial pyramid matching (SPM) causes feature ambiguity near dividing lines because it divides an image into different scales in a fixed manner. A new method called soft SPM (sSPM) is proposed in this paper to reduce feature ambiguity. First, an auxiliary area rotating around a dividing line in four orientations is used to correlate the feature relativity. Second, sSPM is performed to combine these four orientations to describe the image. Finally, an optimized multiple kernel learning (MKL) algorithm with three basic kernels for the support vector machine is applied. Specifically, for each level, a suitable kernel is selected to map the data that fall within the corresponding neighbourhood. In addition, a mixed-norm regularization formulation is optimized using MKL to solve the classification problem. The method proposed in this paper performs well when applied to the Caltech 101 and Scene 15 datasets. Experimental results are collected under various conditions. The results of sSPM are improved by nearly 4% compared with the existing experimental results.

KEY WORDS: Features ambiguity, auxiliary areas, sSPM, Optimized MKL

1 INTRODUCTION

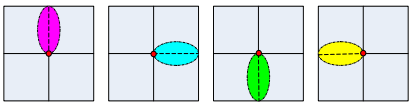
RECENTLY, the combination of spatial pyramid matching (SPM) with the support vector machine (SVM) classifier has been commonly applied to the categorization problem. SPM is an improved version of the bag of words (BoW) model that adds position information by mapping an image into different scales. Traditional dividing methods that use SPM to describe images (Grauman and Darrell, 2005) ignore feature relationships along dividing lines, but features correlated with each other in an image are very useful information for classification. For instance, if there is a computer screen in an office, the mouse and keyboard may be associated with it immediately. If the screen is segmented into two unknown objects, will a viewer still believe that the space is an office? Objects in scenes are always correlated with each other to show the content of an image, whereas boundary features help improve the classification accuracy and object recognition in images.

To reduce the impact of feature ambiguity, an auxiliary area is proposed to correlate feature attributes along dividing lines at each level of SPM. This simple but strong image representation is called soft SPM (sSPM) in this paper. This method is similar to that presented in (Lazebnik, 2006) but has some distinct advantages. With a stronger matching ability than that of SPM and by involving spatial correspondence, more correlated features are observed. At fine levels, correlated features correspond to the original features that fall in the blocks. Specifically, feature in an ambiguous area belongs to which block depends on its distance from the centroid of neighbour blocks. The proposed auxiliary area is shown in four orientations in Figure 1. Different colours represent different feature ambiguity areas. At the coarse level, the number of histograms is not changed when the number of features increases. Therefore, the computational cost of sSPM does not increase. Details concerning which blocks the ambiguity features belong to in sSPM are provided in Section 3.

Kernel selection is introduced for classifiers after the development of an sSPM representation. The SVM is an effective supervised method. It can be used to solve linear or non-linear classification

CONTACT Lu Wu wulv@whut.edu.cn

problems using kernel tricks to map input data into high-dimensional feature spaces. Various kernels are used to measure the similarity between two images. Different kernel candidates are then used. It is a challenge to find an optimal combination of these kernels for specific classification tasks (Bucak, 2014).



(a) 0º (b) 90º (c) 180º (d) 270º

Figure 1. Auxiliary Areas for Soft Assignment

Inspired by Rakotomamonjy (2008) and Varma (2009), the kernel combinations are addressed using mixed-norm regularization functions. This optimal method takes advantage of ℓ_1-norm and ℓ_2-norm regularization. ℓ_1-norm regularization allows prior knowledge to be used in the sSPM calculation, while ℓ_2-norm regularization improves the classification performance by incorporating multiple kernel learning (MKL). Furthermore, sSPM exhibits significant differences for different scales; thus, different single kernel is used to test the performance at each level.

The remainder of this paper is organized as follows. Various related works are presented in Section 2. Section 3 presents the proposed model of sSPM, and an optimal MKL solution method for classification is elaborated. Section 4 experimentally verifies the proposed method using the Caltech 101 and Scene 15 datasets. Finally, Section 5 discusses and concludes the work.

2 RELATED WORKS

MANY works (Yang, 2011; Silva, 2013; Hur, 2015; Yue and Kataqishi, 2016) involved the actual construction of SPM descriptors for classification. Yang (2011) proposed a spatial pyramid co-occurrence method for capturing both the absolute and relative spatial arrangements of words, which characterize a variety of spatial relationships based on the choice and combination of predicates. Silva (2013) proposed incorporating spatial information and the occurrence frequency of visual words to form a graph-based codebook to complete image classification. Hur (2015) refined the existing deformable spatial pyramid model by generalizing the search space and devising spatial smoothness to address the appearance dissimilarities and geometric variations of images. Furthermore, Yue and Kataqishi (2016) utilized the characteristics of sports images and used SPM to obtain absolute feature information and visual word spatial dependence matrices to describe the relative spatial information. Most of these methods incorporated both the absolute and relative spatial information with local descriptors to enhance the image descriptor, whereas the proposed method directly acts on the absolute position to reduce feature ambiguity. The proposed method is simple but effective for image descriptors.

Much work has been performed to combine different features for classification accuracy (Fernando, 2012; Jiang, 2015). Fernando (2012) presented a logistic regression-based fusion method that takes advantage of different global features without being tied to any of them. A marginalized kernel was designed by using the regression model output for image classification. In Jiang (2015), the features of the histogram of oriented gradients (HOG), colour and bar shape were combined with a cell-based histogram structure to form a new HOG-III for human classification and detection. Work regarding learning distances is also relevant to the proposed problem. Guha (2014) proposed a sparse image encoding approach and used the sparsity, quantified based on the compressed distance between the training and query images, for classification. Some details of Euclidean distance function learning were presented by Pan (2006) and Dokmanic (2015). Other feature learning methods, such as Harris-Sift (Zhang, 2012), deep learning framework (Yue, Mao and Li, 2016) and incremental filtering feature selection (Kanimozhi, 2017), are also effective for classification. The techniques are different from the proposed method because the features are not mapped to different scales for computation.

Like feature selection and the integration approach, kernel learning is another method that plays an important role in classification problems. There are many studies regarding the optimization of MKL (Varma and Ray, 2007; Varma and Babu, 2009; Rakotomamonjy, 2008). Varma and Ray (2007) discussed the optimal trade-off for providing a particular training set and prior constraints. Varma and Babu (2009) also extended MKL problems to combine general kernels for different regularization methods. Rakotomamonjy (2008) addressed the MKL problem by using a weighted ℓ2-norm regularization formulation with an additional constraint on the weights that allows sparse kernel combinations. Yan (2014) used adaptive ℓp-norm MKL to learn a robust classifier based on multiple base kernels, which are constructed from concise spatial pyramid features and multiple sets of pre-learned classifiers from other classes. During the kernel learning process, multiple levels of image features are effectively fused, and information is shared among different classifiers. Thiagarajan (2014) proposed performing sparse coding and dictionary learning in the multiple kernel space, where dictionaries are inferred using multiple levels of one-dimensional subspace clustering and sparse codes are obtained using a simple level-wise pursuit scheme. Some researches related to our work; see (Gou, 2014;Tsai, 2014; Ganguly, 2017; Niazmardi,

2018) for examples. The existing research illustrates the importance of kernel learning in the optimization of kernel combinations for specific purposes.

Above all, the existing techniques focus on developing descriptors while ignoring classifier learning for solving specific descriptors. They either focus on improving descriptors or emphasize the theory of object function optimization. In addition, the descriptors of SPM seldom address a single level of different kernels to estimate the performance of different levels. This is very important for designing suitable levels for SPM because model complexity must be considered in the computations. However, the method proposed in this paper considers the problem as a holistic issue. The proposed sSPM method takes advantage of regulated MKL to analyze the performance of each level and to solve the classification problem effectively. This statement is verified in the experimental section.

3 METHOD FOR IMAGE CLASSIFICATION

3.1 Soft Spatial Pyramid Matching

SPM focuses on using the spatial information of features. An image is divided into a number of non-overlapping blocks of the same size. The number of blocks grows exponentially with the number of spatial pyramid levels. The features in one block are always the sum of the features in the four smaller corresponding blocks contained in the next level. Blocks are described in terms of histogram bin counts. Unordered features are individually mapped to multi-resolution histograms. A descriptor is then described as a weighted sum of the histograms of each level.

sSPM refers to soft spatial pyramid matching. It is an extension of SPM and introduces an auxiliary area to enhance the feature correlation. Some useful features are ignored along the dividing line, especially at high levels. An auxiliary area that rotates along the dividing line in four orientations is used to correlate the feature relativity. An image is described using the original blocks in addition to the feature in the auxiliary area. At a coarse level, no auxiliary area is needed, and an image is represented as a BoW. At fine levels, in the same manner as for SPM, an image is divided into 2^{2l} blocks, and 2^{2l} auxiliary areas are established to improve image features expression. The histogram distance is calculated to determine which auxiliary area belongs to which block. The same scheme as that described above is used in the next level. To simplify the computation, the auxiliary area is represented as an ellipse in this paper. Figure 2(a) illustrates the traditional SPM, whereas Figure 2(b) corresponds to sSPM. Both images describe a three-level spatial pyramid. In Figure 2(a), at levels 1 and 2 of the spatial pyramid, the image is individually divided into 4 and 16 blocks. In Figure 2(b), the image is divided in the same manner. However, the difference is that 4 and 16 ellipses are generated for the blocks at levels 1 and 2, respectively. The ellipses are coloured green, pink, yellow and blue in Figure 2(b).

The major axis of an ellipse is used as the board line of the blocks, and the length of the minor axis is set to be a quarter of the major axis. The details are shown in Figure1. Each image is separated into $B=2^{dl}$ bins of dimension d in layer l, and B is the total number of bins. For example, if l is 1, then B equals 4.

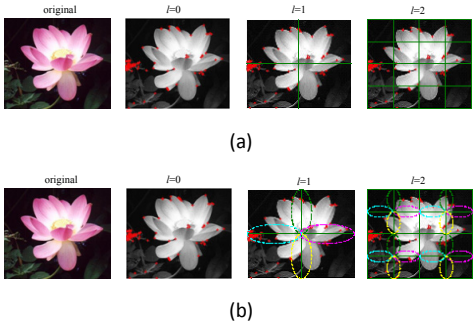


Figure 2. Example of Improved Spatial Pyramid Matching

Suppose that feature X falls in one block, Y falls in the adjacent block and Z falls in an ellipse at each level. Then, H_X, H_Y and H_Z represent histograms of X, Y and Z, respectively. $H_X^l(i)$ and $H_Y^l(i)$ denote the subset of X,Y that falls into the i^{th} bin of level l. The feature similarity between an ellipse and its neighbor blocks is defined by Equation (1).

$$N_l = \left\| H_X^l - H_Z^l \right\|^2 - \left\| H_Y^l - H_Z^l \right\|^2 \quad (1)$$

The value of N_l determines which one is closer to Z. If the value of N_l is a positive value, Z is closer to Y. Then, H_Y is rewritten as H_{Y+Z}. If N_l equals zero, Z is equidistant to both adjacent blocks, and both H_X and H_Y retain their prototypes. From the above calculation, the final descriptor for the image becomes a concatenation of the updated block descriptors.

3.2 Selection of Basic Kernels

Three basic kernels within the SVM classifiers are employed for different levels of sSPM. These basic kernels are proved to be discriminatively powerful and computationally efficient.

Table 1. Three Different Kernels

Type	e.g.	Kernel Function
Linear	Lin.	$< x, y > + c$
Quasi-linear	Poly	$(\alpha < x, y > + c)^d$
Non-linear	RBF	$\exp(-(x-y)^2 / 2\sigma^2)$

The first one is a linear kernel with an inner product $< x, y >$ plus an optional constant c. It is the simplest kernel function and has the lowest complexity compared to that of the other two kernels. The second kernel is a polynomial kernel that is composed of adjustable parameter α, constant c and polynomial degree d. It is also called a quasi-linear kernel. If the degree is 1, it is equivalent to the linear kernel. The third kernel is the widely used radial basis function (RBF) kernel. It has a vital parameter σ, which is adjustable and determines the kernel performance. The complexity of the three kernels increases significantly, as observed in Table 1.

Given the three basic kernels, an optimal linear combination for the SVM classifier is as expressed in Equation (2).

$$K_l = \sum_{m=1}^{M} d_m k_m^l \tag{2}$$

The weights d_m (m=1,2,3) correspond to the trade-off at level l and satisfy $d_m \geq 0, \sum d_m = 1$. The different combinations of d_m lead to the performance variance of the classifiers, and the constraints prevent over- fitting if many basic kernels are involved but only a few are used. M is the total number of kernels and is set to 3 in this paper. The weights learning problem is analyzed in Section 3.3.

3.3 Optimization of MKL

Inspired by Rakotomamonjy (2008) and Varma (2009), the convex optimization problem is described by Equation (3), which is a non-linear objective function with constraints on the simplex.

$$\begin{aligned}
&\min_w \frac{1}{2}\sum_{m=1}^{M}\frac{1}{d_m}\|w_m\|_2^2 + C\sum_{i=1}^{N}\xi_i + \sum_{m=1}^{M} d_m\rho \\
&s.t. y_i(\sum_{m=1}^{M} < w_m, \varphi(x_i^m) > + b) \geq 1 - \xi_i, \\
&\xi_i > 0, i = 1, 2, ...N \\
&\text{where } \varphi < x_i, x_j > = \sum_m d_m \varphi_m^t(x_i)\varphi_m(x_j), \\
&d_m \geq 0, \sum d_m = 1
\end{aligned} \tag{3}$$

This function consists of mixed-norm regularization terms that regularize hyper-planes and kernel combination weights. The first part of Equation (3), which contains d_m, controls the squared norm of the decision function and leads to a much more efficient convex problem. The inner product $< w, w >$ is minimized by increasing the weights and letting the support vectors tend to zero. However, the ℓ_2-norm regularization on d_m will never decrease the performance of larger sets of candidate kernels. In addition, the optimization of this objective function can benefit sparse feature selection. The second d_m, multiplied by ρ, manifests prior information in this object function. The parameter ρ, which encodes prior preferences for descriptors, prevents the weights from becoming too large. Finally, this optimal function takes advantage of both ℓ_1-norm and ℓ_2-norm regularization to increase the flexibility of sSPM. The ℓ_1-norm improves model performance when a small number of kernels is used but degrades the performance when many kernels are combined. However, the ℓ_2-norm regularization improves the model performance regardless of the number of kernels. Thus, the adoption of mixed-norm regulation makes the model flexible and efficient.

To solve the optimal problem in Equation (3), a min-max optimization strategy is adopted. The objective function is reformulated as Equation (4).

$$\begin{aligned}
&T(d) = \min_w \frac{1}{2}\sum_{m=1}^{M}\frac{1}{d_m}\|w_m\|_2^2 + C\sum_{i=1}^{N}\xi_i + \sum_{m=1}^{M} d_m\rho \\
&s.t. y_i(\sum_{m=1}^{M} < w_m, \varphi(x_i^m) > + b) \geq 1 - \xi_i, \\
&\xi_i > 0, i = 1, 2, ...N
\end{aligned} \tag{4}$$

The next step is to solve $T(d)$ using the projected gradient descent via an iterative method. According to the strong duality principle, when α is equal to α*, the function $T(d)$ is equivalent to equation $W(d)$, which is defined by Equation (5).

$$\begin{aligned}
&W(d) = \max_\alpha \sum_{i=1}^{N}\alpha_i^* + \rho^t d - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i^*\alpha_j^* y_i y_j \sum_{m=1}^{M} d_m k_m(x_i, y_i) \\
&s.t.\ \alpha_i \geq 0,\ \sum_{i=1}^{N}\alpha_i y_i = 0.\ \ i, j = 1, 2, ...\text{N}
\end{aligned} \tag{5}$$

where α is the Lagrange multiplier, y denotes the classes and $w = \sum_{i=1}^{n}\alpha_i y_i x_i$. The derivatives of $W(d)$ can be computed if α* does not depend on d_m . Thus, the dual function is differentiated with respect to d_m .

$$\frac{\partial W}{\partial d_m} = \rho_m - \frac{1}{2}\sum_{i,j}\alpha_i^*\alpha_j^* y_i y_j k_m(x_i, y_i) \tag{6}$$

The holistic MKL problem is solved using a two-step iterative method. Both the coefficient α_i and the combination weights d_m are calculated in the optimization of MKL. In the inner loop, a canonical SVM solver is used to calculate α_i at each step with a fixed kernel and given d_m . In the outer loop, d_m is updated using a gradient calculated using the value of α_i found in the inner loop. This two-step iterative method is repeated until convergence.

3.4 d Learning

The multiple kernel weights d_m used in this paper are solved using a gradient descent method and a backtracking line search algorithm. When the gradient of $W(d)$ is obtained, d is updated using a general iteration scheme $d \leftarrow d + \Upsilon D$, where Υ is the step size. As detailed in Table 2, the initial weight is set according to the number of levels in the SPM. To observe the objective function effectively, the step size Υ is determined using a backtracking line search algorithm. After the decent direction D is computed, the first step size is assigned a large positive value. If the objective function decreases, the step size decreases correspondingly. The object function is optimized using the Karush-Kuhn-Tucker (KKT) condition to ensure global convergence.

Table 2. Algorithm of *d* Learning

1.	set $d_m = 1/M$ for $m = 1, 2, ...M$
2.	**while** the KKT condition is not met
3.	**do**
4.	Compute $W(d)$ by using the SVM solver with $K = \sum_m d_m k_m$
5.	Compute $\partial W / \partial d_m$ and a descent direction $D = -\nabla W$
6.	Perform a backtracking line search along D for Υ {calls an SVM solver for the Υ trial value }
7.	$d_m = d_m + \Upsilon D$
8.	**End while**

4 EXPERIMENTS

THE optimal MKL method with sSPM descriptors is tested using synthetic data and the Caltech 101 and Scene 15 datasets. Except for the synthetic data case, the method presented in this paper is also compared with those that use only one descriptor, an enhanced descriptor or non-linear kernels. All the experiments are implemented with Matlab 2012(a).

4.1 Synthetic Data

The proposed method is tested using synthetic data to analyze the impact of different kernels on the movement of support vectors and the variance of the decision line. All kernels from Table 1 take the simple form specified by parameters $c = 0, \alpha = 1, d = 2$ and $\sigma = 1$.

Figure 3 shows the results of applying different kernels to the synthetic data. The support vectors are denoted by red points along the decision line. The red points located far from the decision line degrade the classification accuracy. In Figure 3(a) and (b), the support vectors are more scattered, whereas in (c) and (d), they are closer to the decision line. These results occur because different kernel functions are used: (a), (b) and (c) involve an RBF kernel, a polynomial kernel and a linear kernel, respectively, whereas (d) involves a combination of the above three kernels and thus exhibits the best performance. Obviously, multiple kernels perform better than a single kernel for this synthetic data set.

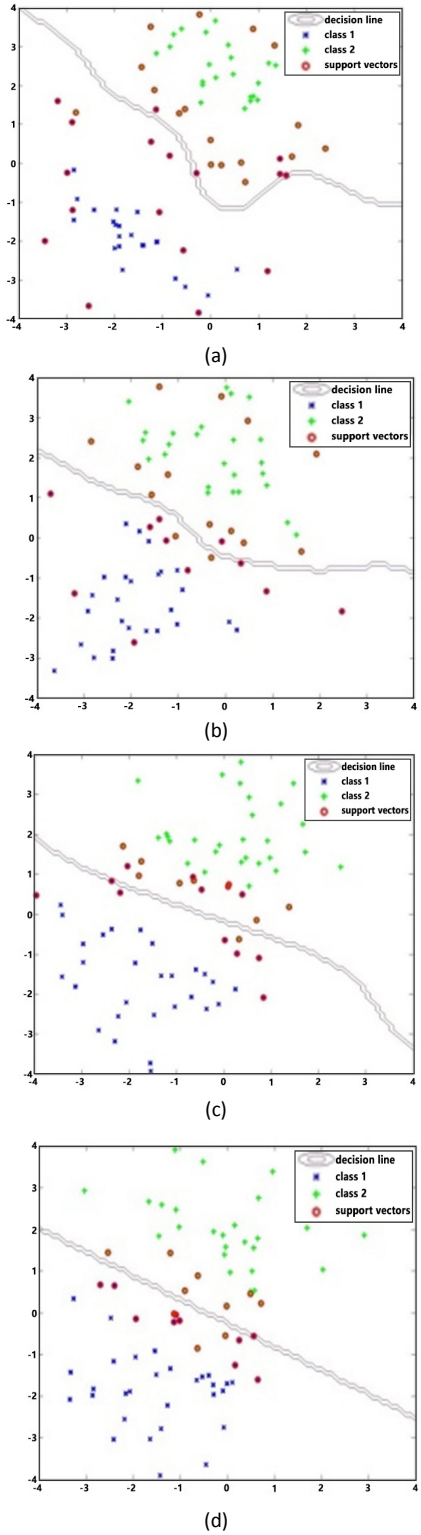


Figure 3. Two-Class Classification of Synthetic Data

4.2 Caltech101 Object Categorization

The experiments test the classification of Caltech 101 from different aspects. Dense SIFT features are extracted at three levels of sSPM. Features in each block are quantified according to a 300-visual-words dictionary using K-means. 30 images are extracted for training, and the remaining images are used for testing in each class. Each image is described by a 6300-dimensional histogram of visual words with 1*1, 2*2 and 4*4 spatial subdivisions. Regarding the parameters of the basic kernels, the constant c is equal to 0 for both linear and polynomial kernels. Furthermore, $\alpha = 1, d = 2$ are set for polynomial kernels, while $\sigma = 1$ is set for the RBF kernel. The MKL is composed of three basic kernels with the same coefficient of 1/3. The prior parameter ρ is set to 0 in this test. To treat multiclass problems, the 1-vs-all formulation is adopted. The results are evaluated in terms of the average accuracy obtained across the classes.

4.2.1 Impact on Different Level Weights of sSPM for Different Kernels

Table 3 shows the different weight combinations of sSPM for the basic kernels and MKL for the classification problem. [1, 0, 0] represents $l_1 = 1, l_2 = 0, l_3 = 0$, which is a BoW. The BoW ignores the spatial position information of features; thus, the results are inferior to those of sSPM. The results of sSPM are at least 4% better than those of the BoW for any basic kernel. This table also shows that sSPM with a single level, i.e. l_2, performs well (71.02%). However, this method is not generally applicable to other datasets. For MKL, the performance is nearly identical to that of linear kernels because linear kernels limit the discriminative ability of other kernels when the feature dimension increases.

Table 3. Different Weight Combinations for Different Kernels (%)

	RBF	Poly	Linear	MKL
[1, 0, 0]	37.73	51.21	43.50	50.90
[0, 1, 0]	**44.13**	**65.93**	**71.02**	**71.02**
[0, 0, 1]	40.98	61.75	59.26	63.16
[1/2, 1/2, 0]	41.98	65.51	70.89	70.89
[1/3, 1/3, 1/3]	42.16	64.97	70.45	70.36
[1/2, 1/4, 1/4]	40.16	64.97	70.45	70.36
[1/4, 1/4, 1/2]	43.22	65.26	70.72	70.72
[1/4, 1/2, 1/4]	44.57	65.38	70.74	70.74

4.2.2 Comparison of Different Numbers of Training Images

Table 4 compares similar methods. The number of training images is increased from 5 to 30 in each class for the entire dataset. MKL is used for the SVM classifier with three-level basic kernels. For the same number of training images, the classification rate is 70.74%, which is better than the rate of 64.6% obtained by Lazebnik (2006). Zhang (2006) used the nearest-neighbour rule to find the minimum distance between the training and query images and obtained a classification accuracy of 66.2%, which is less than that of the proposed method.

Table 4. Different Numbers of Training Images (%)

	5	10	15	20	25	30
SPM (Lazebnik ,2006)	---	---	56.4	---	---	64.60
SVM-KNN(Zhang,2006)	46.60	55.80	59.10	62.00	---	66.20
sSPM	**44.16**	**54.79**	**61.16**	**63.79**	**66.80**	**70.74**

4.2.3 Analysis of Confusion Matrix

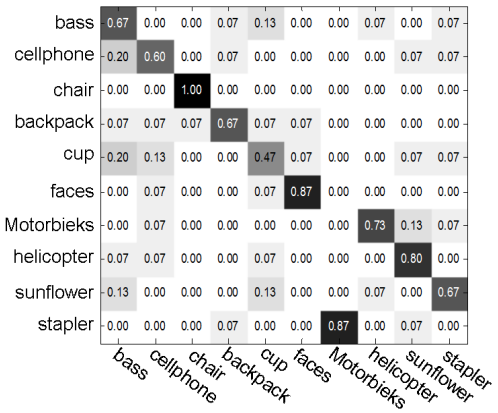
Figure 4 shows the confusion matrices of SPM and sSPM. Caltech 101 contains too many classes; thus, only ten classes are randomly selected to determine what occurs in the inner class. 30 images are also used for training; the remaining images are used for testing in each class. The average accuracy is 41.36% in Figure 4(a) and 43% in Figure 4(b). The optimized MKL produced a 100% classification rate for chairs when SPM was used, but the rate decreased to 93% when sSPM was applied to the same class. However, the accuracy for sunflower increased 6% when sSPM was used. Backpack, faces and sunflowers mutually interfere with each other. In both (a) and (b), the motorbike was difficult to classify because of its complex properties.

4.3 Scene15 Classification

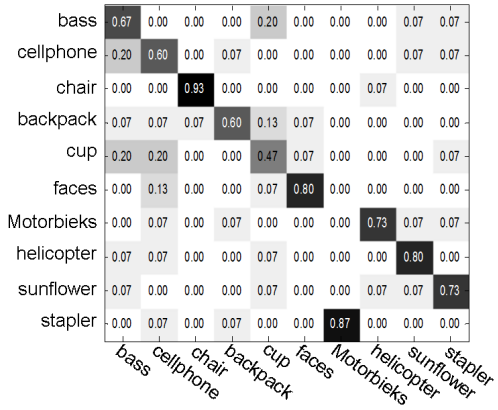
This dataset contains 15 natural scene categories that expand on the thirteen category dataset released by Li (2005). The two new categories are industrial scene and store. The parameters of this experiment are identical to those for the Caltech 101 experiment. Dense SIFT was used to extract and describe the images. In each class, 30 images were used for training, and the remainder were used for testing. The classification results for the fifteen categories are presented in the following subsections.

4.3.1 Impact on Different Level Weights of sSPM for Different Kernels

Table 5 shows that different weight combinations for sSPM yield different results for different kernels. The vector [1, 0, 0] indicates that sSPM can be considered as a BoW model with $l_1 = 1, l_2 = 0$ and $l_3 = 0$. The linear kernel obtains better accuracy than that of the other two single kernels. [0, 1, 0] and [0, 0, 1] represent the single levels l_2 and l_3, respectively. The results show that the descriptor on l_2 and l_3 has



(a) SPM



(b) sSPM

Figure 4. Confusion Matrices of 10 Classes

the same performance for the same kernels, of which the linear kernel also outperforms the other two kernels. In addition, when the weights at levels l_2 and l_3 are equal, the accuracies of the polynomial and linear kernels are identical, while that of RBF is different. These results show that a non-linear kernel is necessary when the data are not separable under linear conditions. In the MKL column, the notable classification performance is presented. According to the last row of Table 5, when empirical weights integrated with the MKL methods were used, sSPM demonstrated the best performance throughout the entire experiment.

Table 5. Different Weight Combinations for Different Kernels (%)

	RBF	Poly	Linear	MKL
[1, 0, 0]	61.12	62.83	68.72	68.77
[0, 1, 0]	64.12	61.78	67.86	67.86
[0, 0, 1]	64.12	61.78	67.86	67.86
[1/3, 1/3, 1/3]	64.86	62.21	68.40	68.30
[1/2, 1/4, 1/4]	63.57	62.21	68.40	68.30
[1/4, 1/4, 1/2]	**64.93**	**62.83**	**68.72**	**68.77**

4.3.2 Comparison of Different Numbers of Training Images

Table 6 shows the classification rates for different numbers of training images. Both SPM and sSPM are divided into three levels and described by the dense SIFT. The accuracy increases as more training images are used. This result is a characteristic of the discriminative model, the quality of which heavily depends on the number of training images.

Table 6. Different Numbers of Training Images (%)

	10	15	20	25	30
BoW	49.62	54.61	56.67	60.87	61.12
SPM	51.86	56.47	59.59	63.89	64.93
sSPM	**51.88**	**56.87**	**59.81**	**64.21**	**68.77**

4.3.3 Analysis of Confusion Matrix

Figure 5 shows the average accuracy for each class. The classification rates are listed along the diagonal. The entry in the i^{th} row and j^{th} column is the percentage of images from class i that are misidentified as belonging to class j. Confusion occurs mainly with indoor scenes, such as the living room (48%), kitchen (52%) and bedroom (59%) images.

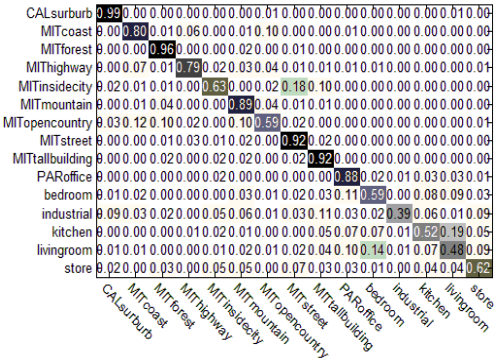


Figure 5. Confusion Matrix of Scene 15 dataset

5 DISCUSSION AND CONCLUSION

IN this paper, sSPM is proposed for feature representation, and mixed-norm regulation is used for the optimization of kernel combinations. First, sSPM utilizes the spatial information of auxiliary areas along dividing lines to reduce feature ambiguity and prevents the loss of important feature information at different scales. Second, as a single unified kernel-based classifier may not satisfactorily solve the classification problem of sSPM, an optimal three-level MKL for the SVM classifier is adopted to improve the classification performance for different levels of a spatial pyramid. Finally, an experiment was conducted to compare the performances of methods that use a

single kernel and a combination of kernels. The experimental results demonstrate the excellence of sSPM with an optimal combination of kernels and that its performance improved by nearly 4% relative to that of the traditional SPM for the datasets presented.

6 ACKNOWLEDGMENT

THE authors would like to acknowledge funding from the National Science Foundation Committee (NSFC) of China (Grant No.51475347 and No. 51675389) and the contributions of all collaborators within the projects mentioned.

7 REFERENCES

S. B. Bucak, R. Jin, and A. K. Jain, (2014). Multiple kernel learning for a visual object recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1354-1369.

I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, (2015). Euclidean distance matrices: Essential theory, algorithms, and applications. *IEEE Signal Processing Magazine,* 32, 12-30.

B. Fernando, F. Elisa, D. Muselet, and M. Sebban, (2012). Discriminative feature fusion for image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 3434-3441.

S. Ganguly, D. Bhattacharjee, and M. Nasipuri, (2017). Fuzzy matching of edge and curvature based features from range images for 3D face recognition. *Intelligent Automation & Soft Computing*, 23(1), 51-62.

K. Grauman and T. Darrell. (2005). The pyramid match kernel: Discriminative classification with sets of image features. *IEEE International Conference on Computer Vision*, 2, 1458-1465.

Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, (2014). A novel MKL model of integrating LiDAR data and MSI for urban area classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 805–818.

T. Guha and R. K. Ward, (2014). Image similarity using sparse representation and compression distance. *IEEE Transactions on Multimedia*, 16, 980–987.

J. Hur, H. Lim, and A. S. Chui (2015). Generalized deformable spatial pyramid: Geometry-preserving dense correspondence estimation. *IEEE Conference on Computer Vision and Pattern Recognition,* 1392-1400.

Y. Jiang and J. Ma, (2015). Combination features and models for human detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 240-248.

U. Kanimozhi and D. Manjula, (2017). An intelligent incremental filtering feature selection and clustering algorithm for effective classification. *Intelligent Automation & Soft Computing*.

S. Lazebnik, C. Schmid, and J. Ponce, (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *IEEE Conference on Computer Vision and Pattern Recognition*, 2, 2169-2178.

F. Li and P. Pietro (2005). A bayesian hierarchical model for learning natural scene categories. *IEEE Conference on Computer Vision and Pattern Recognition,* 2, 524-531.

S. Niazmardi, S. Homayouni, A. Safari, J. L. Shang, and H. McNairn, (2018). Multiple kernel representation and classification of multivariate satellite-image time series for crop mapping. *International Journal of Remote Sensing*, 39(1), 149-168.

Z. B. Pan, T. Ohmi, and K. Kotani, (2006). An efficient method of constructing L1-Type norm feature to estimate euclidean distance for fast vector quantization. *Intelligent Automation & Soft Computing*, 12(3), 269-274.

A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, (2008). Simple MK*L. Journal of Machine Learning Research,* 9, 2491-2521.

F. B. Silva, S. Goldenstein, S. Tabbone, and D. S. Torres, (2013). Image classification based on bag of visual graphs. 20th *IEEE International Conference on Image Processing*, 4312-4316.

J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, (2014). Multiple kernel sparse representations for supervised and unsupervised learning. *IEEE Transaction on Image Processing*, 23, 1057-7149.

J. T. Tsai, Y. Y. Lin, and H.Y.M. Liao, (2014). Per-cluster ensemble kernel learning for multi-modal image clustering with group-dependent feature selection. *IEEE Transaction on Multimedia,* 16,2229-2241.

M. Varma and D. Ray, (2007). Learning the discriminative power-invariance trade-off. *IEEE International Conference on Computer Vision*, 1-8.

M. Varma and B. R. Babu, (2009). More generality in efficient multiple kernel learning. *International Conference on Machine Learning,* 1065-1072.

S. Yan, X. Xu, D. Xu, and S. Lin, (2015). Image classification with densely sampled image windows and generalized adaptive multiple kernel learning. *IEEE Transactions on Cybernetics*,45, 381-390.

Y. Yang and S. Newsam, (2011). Spatial pyramid co-occurrence for image classification. *IEEE International Conference on Computer Vision*, 1465-1472.

G. Yue and K. Kataqishi, (2016). Improved spatial pyramid matching for sports image classification. *IEEE International Conference on Semantic Computering*, 32-38.

J. Yue, S. J. Mao, and M. Li, 2016). A deep learning framework for hyperspectral image classification using spatial pyramid pooling. *Remote Sensing Letters*, 7(9), 875-884.

H. Zhang, A. C. Berg, M. Maire, and J. Malik, (2006). SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. *IEEE Conference on Computer Vision and Pattern Recognition,* 2, 2126-2136.

H.Q. Zhang, C. Xu, X. Gao, and L. Cao, (2012). An indoor mobile visual localization algorithm based on Harris-Sift. *Intelligent Automation & Soft Computing*, 18(7), 885-897.

8 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

9 NOTES ON CONTRIBUTORS

Lu Wu received her B.S. degree and M.S. degree in electronic engineering from Wuhan University of Technology of China (WHUT), in 2004 and 2007 respectively. She is currently working toward the PhD degree at school of information engineering of WHUT. Her research interests include image classification, topic model learning and non-parametric Bayesian learning in computer vision.

Quan Liu received her Ph.D. degree in mechanical manufacturing and Automation from Wuhan University of Technology of China in 2004. She is currently a professor at school of information engineering of Wuhan University of Technology. She is a Council Member of the Chinese Association of Electromagnetic Compatibility and Hubei Institute of Electronics. Her research interests include nonlinear systems theory and signal processing.

Ping Lou received her M.S. degree and Ph.D. degree from Huazhong University of Science and Technology of China, in 1997 and 2004 respectively, both in mechanical engineering. She is currently a professor at school of information engineering of Wuhan University of Technology. Her research interests include digital and intelligent manufacturing, multi-agent system and pattern recognition.